SDS323 Statistical Learning And Inference
Final Project
Nitrogen Dioxide: Air Pollution And Road Traffic At Alnabru In
Oslo, Norway
Lazar Ilic
EID li948

This project was produced entirely by Lazar Ilic. In a 1 session
600 minute round setting on 2022-05-08.

Introduction

The data are a subsample of 500 observations from a data set
that originate in a study where air pollution at a road is
related to traffic volume and meteorological variables,
collected by the Norwegian Public Roads Administration. The
response variable [column 1] consist of hourly values of the
logarithm of the concentration of NO2 [particles], measured at
Alnabru in Oslo, Norway, between October 2001 and August 2003.
The predictor variables [columns 2 to 8] are the logarithm of
the number of cars per hour, temperature 2 meter above ground
[degree C], wind speed [meters/second], the temperature
difference between 25 and 2 meters above ground [degree C], wind
direction [degrees between 0 and 360], hour of day and day
number from Monday October 1. 2001. Submitted by Magne Aldrin
[magne.aldrin@nr.no]. [28/Jul/04] [19kbytes]

I suppose that I might try to simply predict NO2 from the
measurements of the other variables. Or perhaps try to predict
the number of cars per hour from the other variables. It is not
totally clear to me how costly each of those 2 measurements are
or what the sort of latency involved in such measurements is.
How useful these metrics are for other downstream ones. In any
case a simple prediction task, running validation to measure
model performance, and try to produce a good simple model.
Perhaps run a subset selection algorithm or call the nls and glm
functions in .R too maybe to demonstrate just how much estimated
performance can improve in a variety of settings and discuss the
computational aspects of these optimisations and the perhaps
gradient descents involved in finding the optimal coefficients
to minimise error.

Methods

The 1st natural subtask is to simply ideate a little bit. I had previously done a sort of task kind of like this one. With only 500 observations we aim to produce a relatively simple model, maybe 5 or 6 derivative predictor variables, all with strong validity and t-scores. Here, we consider exponentiating certain variables and re transforming them in other more useful ways. As in last time, I considered the day of the week. This is because I expect a different distribution of car types for weekends and weekdays and the number of cars per hour is a very very low resolution metric. Also one might think that there are non trivial sources of air pollution from roads in certain directions, or other potential sources and so bucketing and modifying based upon wind direction may be useful. A natural physics interpretation of air movement and rate of air transfer and throughput off of the road level has to do with the measured temperature values and the Delta which can signify a difference in air pressure or simply reflect facets of the measurement location which I cannot discern simply from afar in this dataset.

A 2nd subtask may be to consult some of the extant literature on this and similar related topics. Here of course we have been engaged in studying the most fundamental basics of statistics from the book An Introduction to Statistical Learning: With Applications in R. But there are papers on Air Pollution and one can query Google, ArXiV, Google Scholar, etc. with "NO2 prediction", "NO2 statistics", "air pollution prediction", etc.

Artificial Intelligence Accuracy Assessment In NO2 Concentration Forecasting Of Metropolises Air
NO2 Prediction Using Machine Learning Analyses In Google EE
Prediction Of The NO2 Concentration Data In An Urban Area Using Multiple Regression And Neuronal Networks
Estimating Lockdown-Induced European NO2 Changes: Using Satellite And Surface Observations And Air Quality Models
A Novel Method For Regional NO2 Concentration Prediction Using Discrete Wavelet Transform And An Long Short Term Memory Network
Spatialisation And Prediction Of Seasonal NO2 Pollution Due To Climate Change In The Korean Capital Area Through Land Use Regression Modeling
Error Prediction Of Air Quality At Monitouring Stations Using Random Forest In A Total Error Framework
Nitrogen Dioxide Levels Estimated From Land Use Regression Models Several Years Apart And Association With Mortality In A Large Cohort Study

A 3rd subtask is to execute a semi comprehensive Exploratory Data Analysis. This comes at the beginning of pretty much every single strong Kaggle Notebook, Project, and Competition submission. So one can produce plots and think about models say with simple subsets of the predictors first like 1 predictor, 2 predictors, plausible interaction terms and study these in isolation and view plots of residuals as in our earlier weeks to examine potential functions for relations. These are reasonable methods.

The 4th subtask is to actually muck around, get our hands dirty, and produce some models for predicting NO2 from the other predictors. Run Cross Validation and Root Mean Squared Error metric as in class to test out our various models and pick one which seems best from our perspective. Perhaps not quite the most performant will be our centaur bet on the actually most useful, valid, and human legible model. We might be prone to overly hacking a little bit and if we try out too many models, the most complex one which produces a strong performance may not really be the best, it may be due to a sort of statistical overfitting or just getting lucky on some pseudorandom complex bullpoopoo. The code will reveal these techniques but it suffices to say that some of the tricks we pulled in class as well as perhaps nls, glm, and other tricks from web logs on these topics come in handy.

The 5th subtask if I get around to it is to predict the number of cars per hour from the other variables. But otherwise it is rather similar to the 4th subtask.

Conclusion And Future Work

Summary Of Results

We produced a simple linear regression model with an $R^2$ of for the provided logarithm of NO2 concentration variable. This corresponds with a Mean Squared Error of  on an average variable of 0.5323 on 5 inputs. So that seems fairly good.

```
Call:
lm(formula = logNO2 ~ logcars + temp + winds + hour + daysplit,
    data = z)
```

Residuals:

```
     Min        1Q    Median       3Q       Max
-2.09434 -0.30500 -0.00121  0.36370   1.85733
```

Coefficients:

| | Estimate | Std. Error | t value | Pr(>|t|) | |
|---|---|---|---|---|---|
| (Intercept) | 1.368052 | 0.172938 | 7.911 | 1.69e-14 | *** |
| logcars | 0.444086 | 0.027539 | 16.126 | < 2e-16 | *** |
| temp | -0.026068 | 0.003776 | -6.903 | 1.57e-11 | *** |
| winds | -0.147835 | 0.013605 | -10.866 | < 2e-16 | *** |
| hour | -0.013929 | 0.004305 | -3.236 | 0.00129 | ** |
| daysplit | -0.292934 | 0.048788 | -6.004 | 3.73e-09 | *** |

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5323 on 494 degrees of freedom
Multiple R-squared:  0.5022,  Adjusted R-squared:  0.4971
F-statistic: 99.66 on 5 and 494 DF,  p-value: < 2.2e-16

Along the way, a number of interesting sub results and observations were made.

Potential Future Work

More data. More information about measurements, instruments, and their errors. More precise geographic data. Ideas about how to modify models given say upcoming future construction information and projects in the region. When someone pings a city for a permit, how far in advance can we use this to predict something useful? How complex are useful models for mapping out future growth of cities trajectories and downstream air pollution dynamics? Will the city respond with a strong tax to disincentivise and thereby lower air pollution? Will the people respond by trying harder to move to suburbs or higher floor apartments? Study the histories of big cities and maybe learn something, London, Los Angeles, Tokyo, New York City, Beijing, Hong Kong, etc. When the pandemic hit how fast were trading firms to predict a variety of trajectories of a variety of different things and how those downstream effects would impact a variety of sectors especially with respect to mid and low frequency trading? Did Nassim Nicholas Taleb make a ton of money shorting a ton back before the pandemic really erupted but he called and predicted that it would in fact blow up to massive epic proportion?

```
# SDS323 Statistical Learning And Inference
# Final Project
# Lazar Ilic
# EID li948
# Sources Referenced: 2 Github R Tutorials, Blogs, and R Documentation

library(tidyverse)
library(tm)
library(ISLR)
#library(ISLR2)
library(e1071)
library(pls)
library(MASS)
library(class)
library(glmnet)
library(leaps)
library(caTools)
library(splines)
library(boot)
library(mgcv)
library(tree)
library(randomForest)
library(gbm)

z=read.csv("../input/inferencedataset/Final.csv")

# "logNO2","logcars","temp","winds","tempd","windd","hour","day"

View(z)
summary(z)

plot(z$logcars,z$logNO2,xlab="logcars",ylab="logNO2",main="logNO2 Versus logcars")
plot(z$temp,z$logNO2,xlab="temp",ylab="logNO2",main="logNO2 Versus temp")
plot(z$winds,z$logNO2,xlab="winds",ylab="logNO2",main="logNO2 Versus winds")
plot(z$tempd,z$logNO2,xlab="tempd",ylab="logNO2",main="logNO2 Versus tempd")
plot(z$windd,z$logNO2,xlab="windd",ylab="logNO2",main="logNO2 Versus windd")
plot(z$hour,z$logNO2,xlab="hour",ylab="logNO2",main="logNO2 Versus hour")
plot((z$day)%%7,z$logNO2,xlab="day",ylab="logNO2",main="logNO2 Versus day")
plot((z$day)%%365,z$logNO2,xlab="day365",ylab="logNO2",main="logNO2 Versus day365")

# It immediately strikes the viewer that logcars looks to be a relatively useful predictor

# "logNO2","logcars","temp","winds","tempd","windd","hour","day"

z$wind1=ifelse(140<=z$windd & z$windd<=360,1,0)
#View(z)
z$daytime=ifelse(8<=z$hour & z$hour<=18,1,0)
z$day7=(z$day)%%7
z$day365=(z$day)%%365
z$daysplit=ifelse(z$day7<1.5 | z$day7>5.5,1,0)
```

```
View(z)

# Do a little bit of the old un hacked eyeballed binary Boolean one hot dummy variable cor

# "logNO2","logcars","temp","winds","tempd","windd","hour","day", "wind1", "daytime", "day

Model.logcars=lm(logNO2 ~ logcars,z)
summary(Model.logcars)
plot(Model.logcars$fitted.values,Model.logcars$residuals,main="Residuals For Model.logcars
hist(Model.logcars$residuals)

Model.temp=lm(logNO2 ~ temp,z)
summary(Model.temp)
plot(Model.temp$fitted.values,Model.temp$residuals,main="Residuals For Model.temp",ylab="F
hist(Model.temp$residuals)

Model.winds=lm(logNO2 ~ winds,z)
summary(Model.winds)
plot(Model.winds$fitted.values,Model.winds$residuals,main="Residuals For Model.winds",ylab
hist(Model.winds$residuals)

Model.tempd=lm(logNO2 ~ tempd,z)
summary(Model.tempd)
plot(Model.tempd$fitted.values,Model.tempd$residuals,main="Residuals For Model.tempd",ylab
hist(Model.tempd$residuals)

Model.windd=lm(logNO2 ~ windd,z)
summary(Model.windd)
plot(Model.windd$fitted.values,Model.windd$residuals,main="Residuals For Model.windd",ylab
hist(Model.windd$residuals)

# "logNO2","logcars","temp","winds","tempd","windd","hour","day", "wind1", "daytime", "day

Model.hour=lm(logNO2 ~ hour,z)
summary(Model.hour)
plot(Model.hour$fitted.values,Model.hour$residuals,main="Residuals For Model.hour",ylab="F
hist(Model.hour$residuals)

Model.wind1=lm(logNO2 ~ wind1,z)
summary(Model.wind1)
plot(Model.wind1$fitted.values,Model.wind1$residuals,main="Residuals For Model.wind1",ylab
hist(Model.wind1$residuals)

Model.daytime=lm(logNO2 ~ daytime,z)
summary(Model.daytime)
plot(Model.daytime$fitted.values,Model.daytime$residuals,main="Residuals For Model.daytime
hist(Model.daytime$residuals)

Model.day7=lm(logNO2 ~ day7,z)
summary(Model.day7)
plot(Model.day7$fitted.values,Model.day7$residuals,main="Residuals For Model.day7",ylab="F
```

```r
hist(Model.day7$residuals)

Model.day365=lm(logNO2 ~ day365,z)
summary(Model.day365)
plot(Model.day365$fitted.values,Model.day365$residuals,main="Residuals For Model.day365",y
hist(Model.day365$residuals)

Model.daysplit=lm(logNO2 ~ daysplit,z)
summary(Model.daysplit)
plot(Model.daysplit$fitted.values,Model.daysplit$residuals,main="Residuals For Model.daysp
hist(Model.daysplit$residuals)

# "logNO2","logcars","temp","winds","tempd","windd","hour","day","wind1","daytime","day7",

# logcars seems useful. temp seems useful. winds seems useful. hour seems useful. daytime

Model1=lm(logNO2 ~ logcars + temp + winds + hour + daytime + daysplit,z)
summary(Model1)
plot(Model1$fitted.values,Model1$residuals,main="Residuals For Model1",ylab="Residual",xla
hist(Model1$residuals)

# OK now a huge loss of significance on daytime.

Model2=lm(logNO2 ~ logcars + temp + winds + hour + daysplit,z)
summary(Model2)
plot(Model2$fitted.values,Model2$residuals,main="Residuals For Model2",ylab="Residual",xla
hist(Model2$residuals)

# Could try more sophisticated things like glm, nls. Or forming more complex models and ru
```

```
── Attaching packages ──────────────────────────────────────── tidyverse 1.3.1 ──

✔ ggplot2 3.3.5      ✔ purrr   0.3.4
✔ tibble  3.1.5      ✔ dplyr   1.0.7
✔ tidyr   1.1.4      ✔ stringr 1.4.0
✔ readr   2.0.2      ✔ forcats 0.5.1
── Conflicts ─────────────────────────────────────────── tidyverse_conflicts() ──
✖ dplyr::filter() masks stats::filter()
✖ dplyr::lag()    masks stats::lag()


Loading required package: NLP


Attaching package: 'NLP'


The following object is masked from 'package:ggplot2':

    annotate


The following object is masked from 'package:httr':

    content


Attaching package: 'pls'


The following object is masked from 'package:stats':

    loadings


Attaching package: 'MASS'


The following object is masked from 'package:dplyr':

    select


Loading required package: Matrix


Attaching package: 'Matrix'


The following objects are masked from 'package:tidyr':
```

```
     expand, pack, unpack



Loaded glmnet 4.1-1

Loading required package: nlme


Attaching package: 'nlme'


The following object is masked from 'package:dplyr':

     collapse


This is mgcv 1.8-35. For overview type 'help("mgcv-package")'.

Registered S3 method overwritten by 'tree':
  method     from
  print.tree cli

randomForest 4.6-14

Type rfNews() to see new features/changes/bug fixes.


Attaching package: 'randomForest'


The following object is masked from 'package:dplyr':

     combine


The following object is masked from 'package:ggplot2':

     margin


Loaded gbm 2.1.8
```

A data.frame: 500 × 8

| logNO2 | logcars | temp | winds | tempd | windd | hour | day |
|--------|---------|------|-------|-------|-------|------|-----|
| <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | <int> | <int> |
| 3.71844 | 7.69120 | 9.2 | 4.8 | -0.1 | 74.4 | 20 | 600 |
| 3.10009 | 7.69894 | 6.4 | 3.5 | -0.3 | 56.0 | 14 | 196 |
| 3.31419 | 4.81218 | -3.7 | 0.9 | -0.1 | 281.3 | 4 | 513 |
| 4.38826 | 6.95177 | -7.2 | 1.7 | 1.2 | 74.0 | 23 | 143 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 4.34640 | 7.51806 | -1.3 | 2.6 | -0.1 | 65.0 | 11 | 115 |
| 4.16044 | 7.67183 | 2.6 | 1.6 | 0.3 | 224.2 | 19 | 527 |
| 4.01277 | 5.52545 | -7.9 | 1.6 | 0.3 | 211.9 | 5 | 502 |
| 2.15176 | 4.68213 | -4.1 | 3.8 | -0.1 | 63.1 | 4 | 453 |
| 3.15700 | 7.15618 | -12.7 | 5.2 | -0.1 | 64.5 | 12 | 462 |
| 2.37955 | 4.74493 | -1.6 | 3.0 | 0.4 | 58.3 | 3 | 554 |
| 3.83298 | 5.81114 | -3.1 | 1.8 | 0.3 | 78.0 | 2 | 55 |
| 4.48187 | 8.10892 | 1.0 | 1.2 | 1.5 | 215.0 | 18 | 47 |
| 4.04830 | 8.31385 | 12.2 | 4.0 | -2.8 | 230.4 | 17 | 572 |
| 4.00186 | 5.22036 | -1.5 | 2.4 | 0.9 | 82.7 | 5 | 556 |
| 3.23080 | 6.40853 | -0.9 | 3.0 | 0.1 | 235.0 | 1 | 69 |
| 4.67189 | 7.31920 | -8.5 | 0.8 | 2.9 | 282.4 | 20 | 447 |
| 2.73437 | 6.61740 | 6.5 | 4.1 | 0.0 | 88.0 | 24 | 186 |
| 3.49651 | 7.76938 | 4.2 | 7.1 | -1.1 | 9.3 | 10 | 550 |
| 3.67122 | 6.46770 | -1.1 | 2.5 | 1.0 | 277.0 | 6 | 142 |
| 3.67377 | 7.65064 | 0.8 | 3.4 | -0.4 | 70.0 | 12 | 167 |
| 3.15274 | 7.75061 | 8.2 | 4.5 | 0.2 | 307.0 | 14 | 32 |
| 3.42751 | 5.18178 | 0.2 | 0.4 | 0.2 | 230.0 | 5 | 112 |
| 4.32413 | 7.63964 | -2.1 | 4.3 | -0.2 | 41.4 | 11 | 432 |
| 3.65584 | 8.00703 | -2.8 | 6.0 | -0.2 | 62.5 | 15 | 453 |
| 3.93183 | 7.75319 | -5.4 | 2.5 | 1.4 | 83.0 | 20 | 98 |
| 2.28238 | 4.65396 | -5.7 | 4.9 | -0.1 | 58.5 | 4 | 409 |
| 4.13035 | 6.67582 | 6.6 | 1.6 | 0.4 | 104.5 | 23 | 579 |
| 3.77963 | 7.79688 | 5.7 | 4.8 | -1.5 | 223.0 | 11 | 178 |
| 3.97968 | 8.02224 | 13.8 | 0.8 | -0.1 | 201.6 | 17 | 602 |
| 3.75420 | 7.48829 | 3.2 | 4.2 | -0.6 | 285.0 | 12 | 121 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 2.98062 | 6.89163 | 0.1 | 1.0 | 0.5 | 334.0 | 23 | 66 |
| 3.65066 | 7.97247 | -4.8 | 3.7 | 0.0 | 16.0 | 18 | 116 |
| 4.25135 | 7.75018 | 9.9 | 1.8 | -1.3 | 266.3 | 14 | 529 |
| 3.65842 | 7.42536 | 4.0 | 5.8 | 0.2 | 184.6 | 20 | 402 |
| 3.99083 | 7.63723 | -5.4 | 4.1 | -0.2 | 72.5 | 17 | 440 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 4.44147 | 6.55108 | -12.0 | 2.8 | 0.4 | 82.5 | 6 | 444 |
| 2.28238 | 4.98361 | 1.8 | 3.5 | 0.3 | 121.0 | 4 | 128 |
| 3.92790 | 6.57508 | 5.0 | 0.8 | 0.3 | 262.7 | 6 | 403 |
| 3.35690 | 7.09257 | 4.6 | 3.4 | 0.2 | 194.0 | 22 | 109 |
| 4.18814 | 7.83201 | -0.2 | 0.6 | -0.3 | 99.7 | 7 | 541 |
| 4.20170 | 7.55329 | -0.9 | 1.1 | -0.3 | 92.0 | 13 | 51 |
| 2.41591 | 7.23778 | 11.9 | 4.2 | -1.3 | 233.0 | 15 | 180 |
| 4.33336 | 8.20084 | -9.0 | 2.4 | 0.0 | 84.0 | 8 | 100 |
| 1.58924 | 5.81711 | 1.3 | 3.8 | 0.3 | 33.0 | 5 | 174 |
| 4.04830 | 8.26462 | 7.7 | 2.6 | -0.8 | 69.1 | 16 | 571 |
| 2.65324 | 5.05625 | -7.0 | 3.2 | -0.1 | 67.0 | 20 | 85 |
| 4.06732 | 7.73193 | 2.6 | 4.2 | -0.2 | 77.9 | 14 | 514 |
| 3.86283 | 5.99396 | 1.6 | 1.7 | 1.5 | 209.1 | 1 | 545 |
| 3.98155 | 7.64108 | 6.7 | 2.3 | -0.4 | 247.0 | 10 | 33 |
| 3.81331 | 7.74587 | 1.6 | 2.1 | -0.1 | 146.0 | 17 | 167 |
| 4.31615 | 5.14749 | -6.5 | 1.7 | 0.8 | 184.0 | 5 | 154 |
| 3.33932 | 5.66643 | 6.7 | 2.4 | -0.1 | 66.1 | 5 | 592 |
| 3.96651 | 7.34923 | -0.4 | 1.9 | 1.0 | 77.0 | 21 | 72 |
| 3.86073 | 4.45435 | -13.8 | 4.2 | 0.2 | 80.0 | 4 | 95 |
| 2.56495 | 4.58497 | 1.8 | 2.3 | 0.1 | 59.0 | 4 | 164 |
| 4.30946 | 7.68202 | 3.5 | 5.0 | -1.0 | 78.0 | 11 | 166 |
| 2.94444 | 6.52942 | 9.5 | 6.5 | -0.8 | 210.0 | 10 | 35 |
| 4.17439 | 7.75791 | 5.2 | 4.6 | -0.8 | 214.0 | 14 | 176 |
| 2.95491 | 5.78996 | 8.4 | 0.5 | -2.6 | 108.5 | 7 | 588 |
| 4.03247 | 8.16223 | 4.7 | 5.9 | 0.4 | 207.0 | 17 | 128 |

```
     logNO2           logcars           temp               winds
 Min.   :1.224    Min.   :4.127    Min.   :-18.6000    Min.   :0.300
 1st Qu.:3.214    1st Qu.:6.176    1st Qu.: -3.9000    1st Qu.:1.675
 Median :3.848    Median :7.425    Median :  1.1000    Median :2.800
 Mean   :3.698    Mean   :6.973    Mean   :  0.8474    Mean   :3.056
 3rd Qu.:4.217    3rd Qu.:7.793    3rd Qu.:  4.9000    3rd Qu.:4.200
 Max.   :6.395    Max.   :8.349    Max.   : 21.1000    Max.   :9.900
      tempd            windd             hour               day
 Min.   :-5.4000   Min.   :  2.0    Min.   : 1.00    Min.   : 32.0
 1st Qu.:-0.2000   1st Qu.: 72.0    1st Qu.: 6.00    1st Qu.:118.8
 Median : 0.0000   Median : 97.0    Median :12.50    Median :212.0
 Mean   : 0.1494   Mean   :143.4    Mean   :12.38    Mean   :210.5
```

```
Mean    : 0.1494    Mean    :143.4    Mean    :12.38    Mean    :310.5
3rd Qu.: 0.6000    3rd Qu.:220.0    3rd Qu.:18.00    3rd Qu.:513.0
Max.    : 4.3000    Max.    :359.0    Max.    :24.00    Max.    :608.0
```
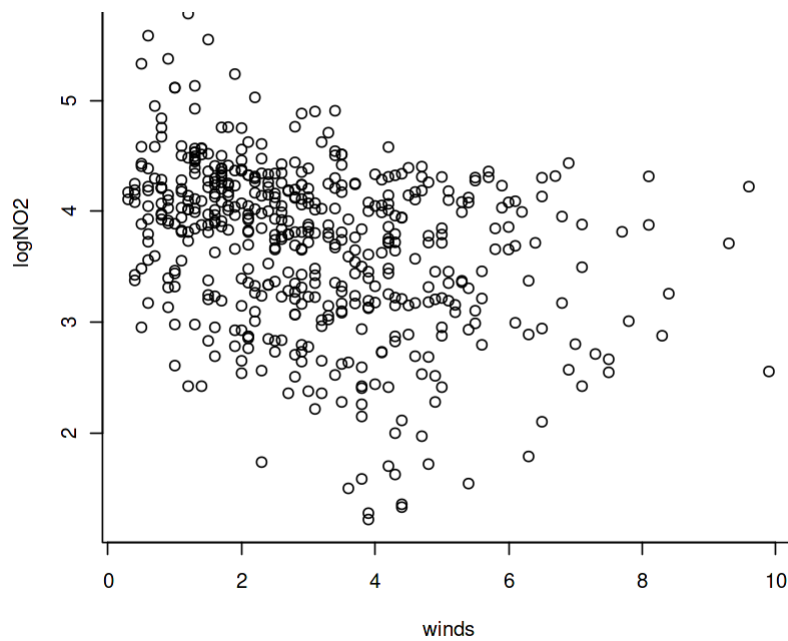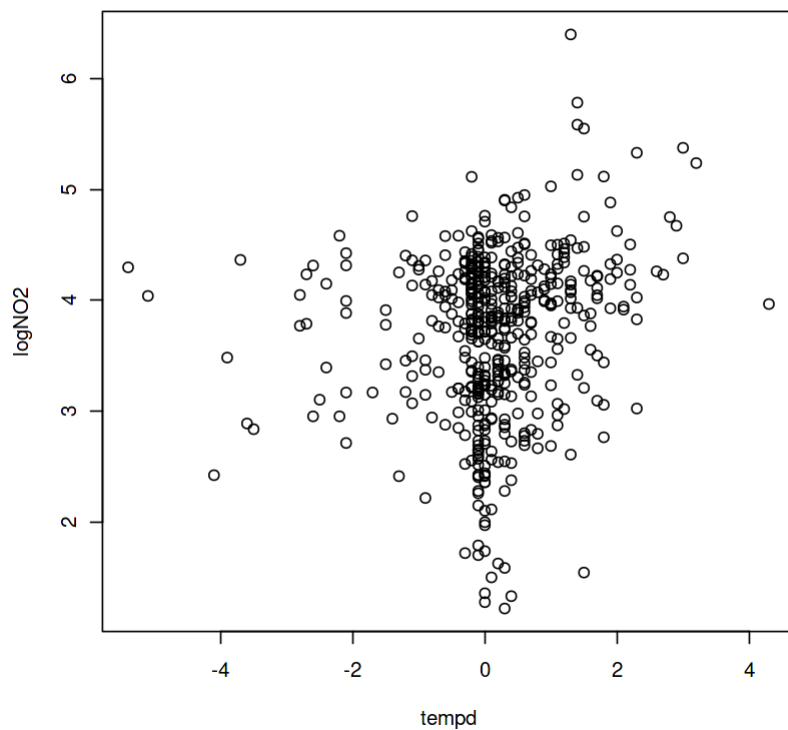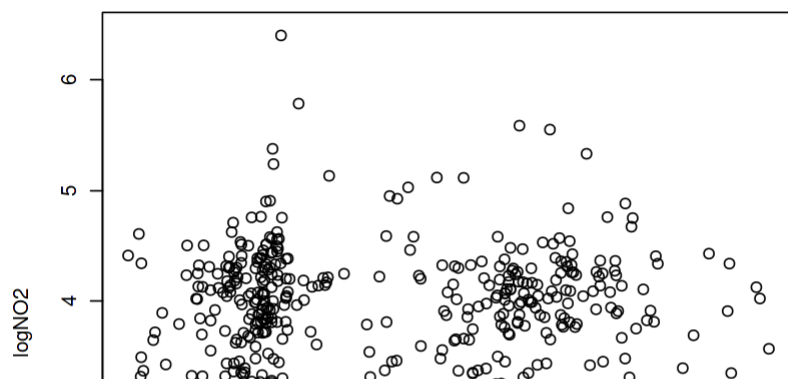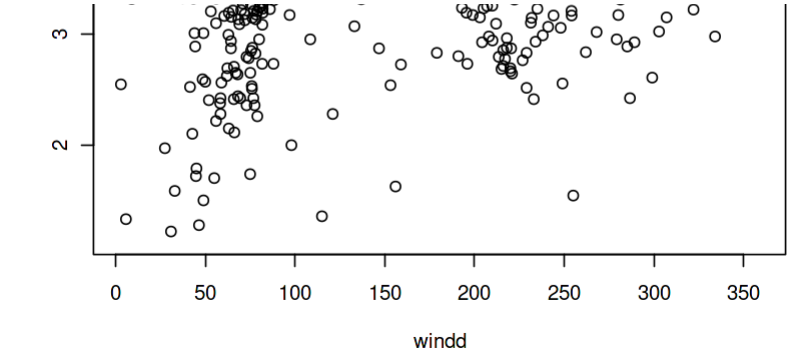
**logNO2 Versus logcars**
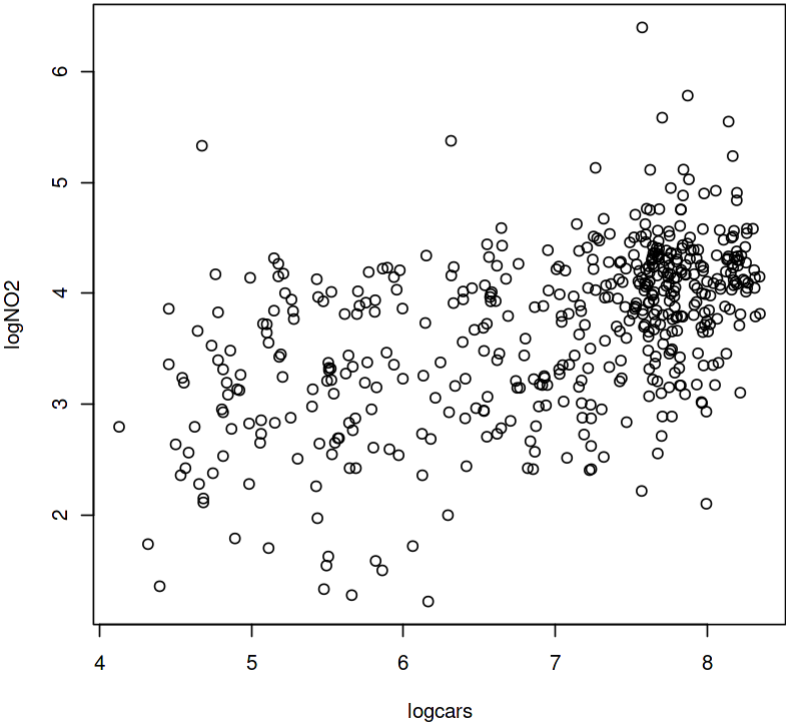


logcars

**logNO2 Versus temp**
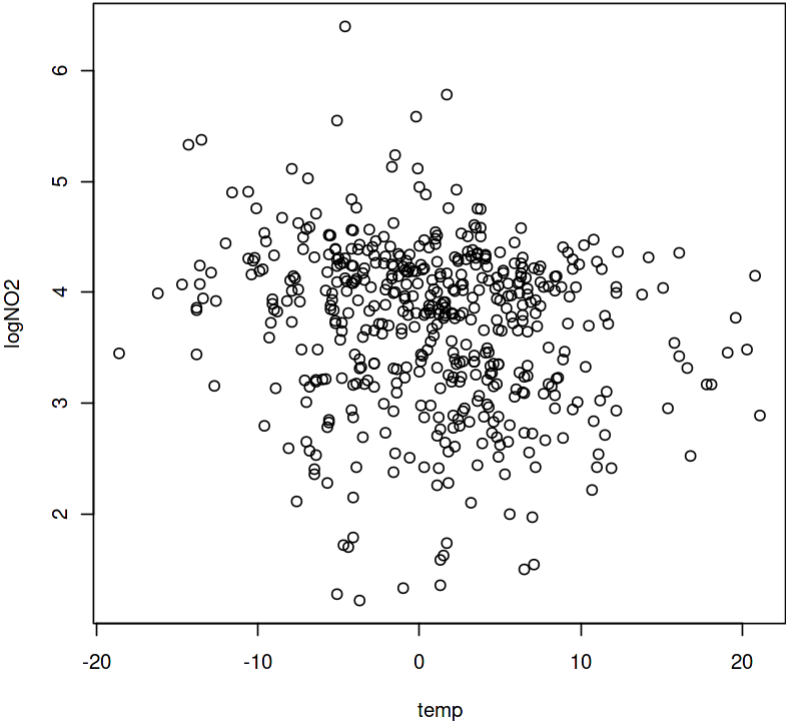


temp

**logNO2 Versus winds**

## logNO2 Versus tempd



## logNO2 Versus windd

logNO2 Versus hour
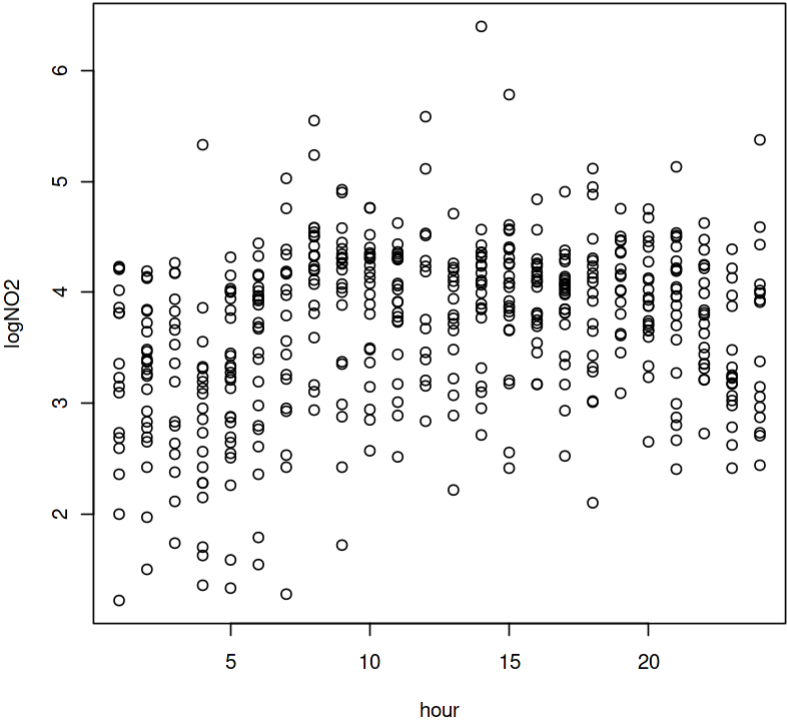


logNO2 Versus day
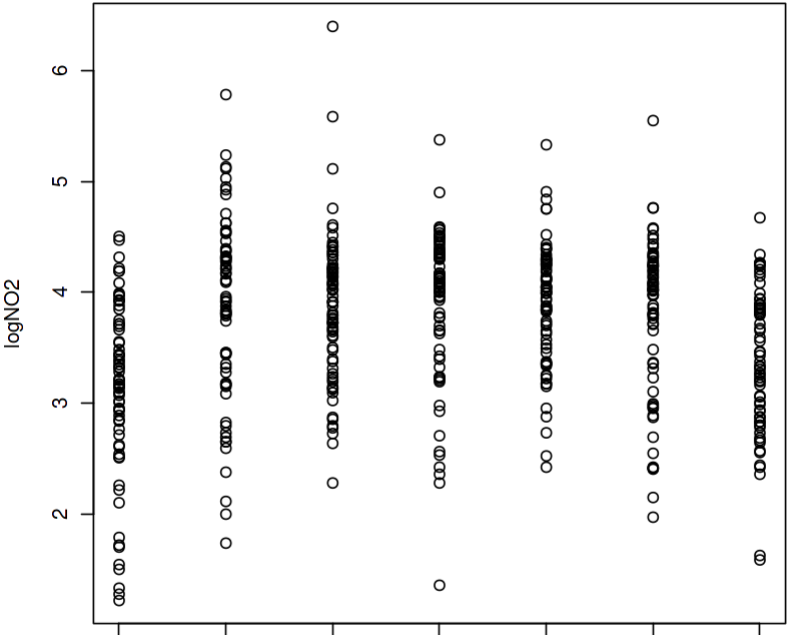
| 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|

day

A data.frame: 500 × 13

| logNO2 | logcars | temp | winds | tempd | windd | hour | day | wind1 | daytime | day7 | day3 |
| <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | <int> | <int> | <dbl> | <dbl> | <dbl> | <dbl> |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 3.71844 | 7.69120 | 9.2 | 4.8 | -0.1 | 74.4 | 20 | 600 | 0 | 0 | 5 | 2 |
| 3.10009 | 7.69894 | 6.4 | 3.5 | -0.3 | 56.0 | 14 | 196 | 0 | 1 | 0 | 1 |
| 3.31419 | 4.81218 | -3.7 | 0.9 | -0.1 | 281.3 | 4 | 513 | 1 | 0 | 2 | 1 |
| 4.38826 | 6.95177 | -7.2 | 1.7 | 1.2 | 74.0 | 23 | 143 | 0 | 0 | 3 | 1 |
| 4.34640 | 7.51806 | -1.3 | 2.6 | -0.1 | 65.0 | 11 | 115 | 0 | 1 | 3 | 1 |
| 4.16044 | 7.67183 | 2.6 | 1.6 | 0.3 | 224.2 | 19 | 527 | 1 | 0 | 2 | 1 |
| 4.01277 | 5.52545 | -7.9 | 1.6 | 0.3 | 211.9 | 5 | 502 | 1 | 0 | 5 | 1 |
| 2.15176 | 4.68213 | -4.1 | 3.8 | -0.1 | 63.1 | 4 | 453 | 0 | 0 | 5 | |
| 3.15700 | 7.15618 | -12.7 | 5.2 | -0.1 | 64.5 | 12 | 462 | 0 | 1 | 0 | |
| 2.37955 | 4.74493 | -1.6 | 3.0 | 0.4 | 58.3 | 3 | 554 | 0 | 0 | 1 | 1 |
| 3.83298 | 5.81114 | -3.1 | 1.8 | 0.3 | 78.0 | 2 | 55 | 0 | 0 | 6 | |
| 4.48187 | 8.10892 | 1.0 | 1.2 | 1.5 | 215.0 | 18 | 47 | 1 | 1 | 5 | |
| 4.04830 | 8.31385 | 12.2 | 4.0 | -2.8 | 230.4 | 17 | 572 | 1 | 1 | 5 | 2 |
| 4.00186 | 5.22036 | -1.5 | 2.4 | 0.9 | 82.7 | 5 | 556 | 0 | 0 | 3 | 1 |
| 3.23080 | 6.40853 | -0.9 | 3.0 | 0.1 | 235.0 | 1 | 69 | 1 | 0 | 6 | |
| 4.67189 | 7.31920 | -8.5 | 0.8 | 2.9 | 282.4 | 20 | 447 | 1 | 0 | 6 | |
| 2.73437 | 6.61740 | 6.5 | 4.1 | 0.0 | 88.0 | 24 | 186 | 0 | 0 | 4 | 1 |
| 3.49651 | 7.76938 | 4.2 | 7.1 | -1.1 | 9.3 | 10 | 550 | 0 | 1 | 4 | |
| 3.67122 | 6.46770 | -1.1 | 2.5 | 1.0 | 277.0 | 6 | 142 | 1 | 0 | 2 | 1 |
| 3.67377 | 7.65064 | 0.8 | 3.4 | -0.4 | 70.0 | 12 | 167 | 0 | 1 | 6 | 1 |
| 3.15274 | 7.75061 | 8.2 | 4.5 | 0.2 | 307.0 | 14 | 32 | 1 | 1 | 4 | |
| 3.42751 | 5.18178 | 0.2 | 0.4 | 0.2 | 230.0 | 5 | 112 | 1 | 0 | 0 | 1 |
| 4.32413 | 7.63964 | -2.1 | 4.3 | -0.2 | 41.4 | 11 | 432 | 0 | 1 | 5 | |
| 3.65584 | 8.00703 | -2.8 | 6.0 | -0.2 | 62.5 | 15 | 453 | 0 | 1 | 5 | |
| 3.93183 | 7.75319 | -5.4 | 2.5 | 1.4 | 83.0 | 20 | 98 | 0 | 0 | 0 | |
| 2.28238 | 4.65396 | -5.7 | 4.9 | -0.1 | 58.5 | 4 | 409 | 0 | 0 | 3 | |
| 4.13035 | 6.67582 | 6.6 | 1.6 | 0.4 | 104.5 | 23 | 579 | 0 | 0 | 5 | 2 |
| 3.77963 | 7.79688 | 5.7 | 4.8 | -1.5 | 223.0 | 11 | 178 | 1 | 1 | 3 | 1 |

| 3.77963 | 7.79688 | 3.7 | 4.8 | 1.8 | 223.0 | 11 | 178 | 1 | 1 | 3 | |
| 3.97968 | 8.02224 | 13.8 | 0.8 | -0.1 | 201.6 | 17 | 602 | 1 | 1 | 0 | 2 |
| 3.75420 | 7.48829 | 3.2 | 4.2 | -0.6 | 285.0 | 12 | 121 | 1 | 1 | 2 | 1 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | |
| 2.98062 | 6.89163 | 0.1 | 1.0 | 0.5 | 334.0 | 23 | 66 | 1 | 0 | 3 | |
| 3.65066 | 7.97247 | -4.8 | 3.7 | 0.0 | 16.0 | 18 | 116 | 0 | 1 | 4 | 1 |
| 4.25135 | 7.75018 | 9.9 | 1.8 | -1.3 | 266.3 | 14 | 529 | 1 | 1 | 4 | 1 |
| 3.65842 | 7.42536 | 4.0 | 5.8 | 0.2 | 184.6 | 20 | 402 | 1 | 0 | 3 | |
| 3.99083 | 7.63723 | -5.4 | 4.1 | -0.2 | 72.5 | 17 | 440 | 0 | 1 | 6 | |
| 4.44147 | 6.55108 | -12.0 | 2.8 | 0.4 | 82.5 | 6 | 444 | 0 | 0 | 3 | |
| 2.28238 | 4.98361 | 1.8 | 3.5 | 0.3 | 121.0 | 4 | 128 | 0 | 0 | 2 | 1 |
| 3.92790 | 6.57508 | 5.0 | 0.8 | 0.3 | 262.7 | 6 | 403 | 1 | 0 | 4 | |
| 3.35690 | 7.09257 | 4.6 | 3.4 | 0.2 | 194.0 | 22 | 109 | 1 | 0 | 4 | 1 |
| 4.18814 | 7.83201 | -0.2 | 0.6 | -0.3 | 99.7 | 7 | 541 | 0 | 0 | 2 | 1 |
| 4.20170 | 7.55329 | -0.9 | 1.1 | -0.3 | 92.0 | 13 | 51 | 0 | 1 | 2 | |
| 2.41591 | 7.23778 | 11.9 | 4.2 | -1.3 | 233.0 | 15 | 180 | 1 | 1 | 5 | 1 |
| 4.33336 | 8.20084 | -9.0 | 2.4 | 0.0 | 84.0 | 8 | 100 | 0 | 1 | 2 | 1 |
| 1.58924 | 5.81711 | 1.3 | 3.8 | 0.3 | 33.0 | 5 | 174 | 0 | 0 | 6 | 1 |
| 4.04830 | 8.26462 | 7.7 | 2.6 | -0.8 | 69.1 | 16 | 571 | 0 | 1 | 4 | 2 |
| 2.65324 | 5.05625 | -7.0 | 3.2 | -0.1 | 67.0 | 20 | 85 | 0 | 0 | 1 | |
| 4.06732 | 7.73193 | 2.6 | 4.2 | -0.2 | 77.9 | 14 | 514 | 0 | 1 | 3 | 1 |
| 3.86283 | 5.99396 | 1.6 | 1.7 | 1.5 | 209.1 | 1 | 545 | 1 | 0 | 6 | 1 |
| 3.98155 | 7.64108 | 6.7 | 2.3 | -0.4 | 247.0 | 10 | 33 | 1 | 1 | 5 | |
| 3.81331 | 7.74587 | 1.6 | 2.1 | -0.1 | 146.0 | 17 | 167 | 1 | 1 | 6 | 1 |
| 4.31615 | 5.14749 | -6.5 | 1.7 | 0.8 | 184.0 | 5 | 154 | 1 | 0 | 0 | 1 |
| 3.33932 | 5.66643 | 6.7 | 2.4 | -0.1 | 66.1 | 5 | 592 | 0 | 0 | 4 | 2 |
| 3.96651 | 7.34923 | -0.4 | 1.9 | 1.0 | 77.0 | 21 | 72 | 0 | 0 | 2 | |
| 3.86073 | 4.45435 | -13.8 | 4.2 | 0.2 | 80.0 | 4 | 95 | 0 | 0 | 4 | |
| 2.56495 | 4.58497 | 1.8 | 2.3 | 0.1 | 59.0 | 4 | 164 | 0 | 0 | 3 | 1 |
| 4.30946 | 7.68202 | 3.5 | 5.0 | -1.0 | 78.0 | 11 | 166 | 0 | 1 | 5 | 1 |
| 2.94444 | 6.52942 | 9.5 | 6.5 | -0.8 | 210.0 | 10 | 35 | 1 | 1 | 0 | |
| 4.17439 | 7.75791 | 5.2 | 4.6 | -0.8 | 214.0 | 14 | 176 | 1 | 1 | 1 | 1 |

| 2.95491 | 5.78996 | 8.4 | 0.5 | -2.6 | 108.5 | 7 | 588 | 0 | 0 | 0 | 2 |
| 4.03247 | 8.16223 | 4.7 | 5.9 | 0.4 | 207.0 | 17 | 128 | 1 | 1 | 2 |  |

```
Call:
lm(formula = logNO2 ~ logcars, data = z)

Residuals:
     Min      1Q  Median      3Q     Max
-2.18822 -0.40071  0.06428  0.40362  2.48472

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.23310    0.18755   6.575 1.23e-10 ***
logcars      0.35353    0.02657  13.303  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6454 on 498 degrees of freedom
Multiple R-squared:  0.2622,    Adjusted R-squared:  0.2607
F-statistic:    177 on 1 and 498 DF,  p-value: < 2.2e-16
```

**logNO2 Versus day365**



**Residuals For Model.logcars**
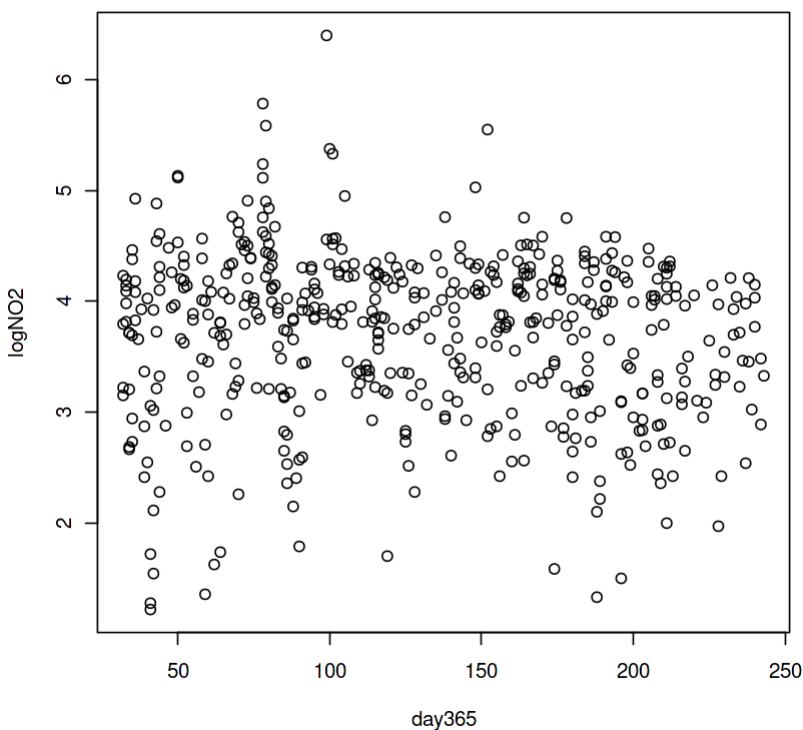
Fitted Values Of Model.logcars
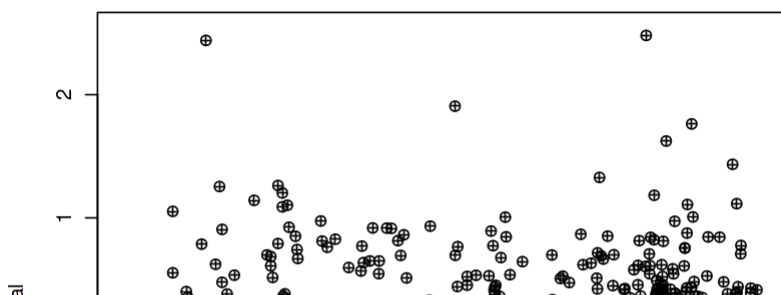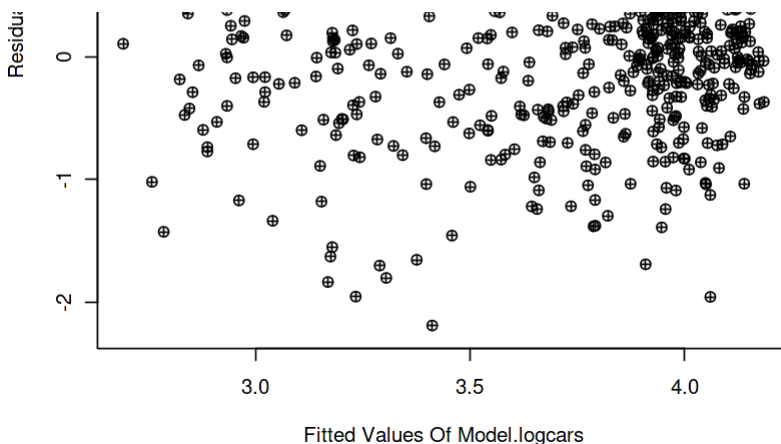
```
Call:
lm(formula = logNO2 ~ temp, data = z)

Residuals:
    Min      1Q  Median      3Q     Max
-2.5626 -0.4509  0.1349  0.5037  2.5913

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.714761   0.033402 111.215  < 2e-16 ***
temp        -0.019345   0.005082  -3.807 0.000158 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7407 on 498 degrees of freedom
Multiple R-squared:  0.02828,   Adjusted R-squared:  0.02633
F-statistic: 14.49 on 1 and 498 DF,  p-value: 0.0001583
```

### Histogram of Model.logcars$residuals



Model.logcars$residuals

**Residuals For Model.temp**
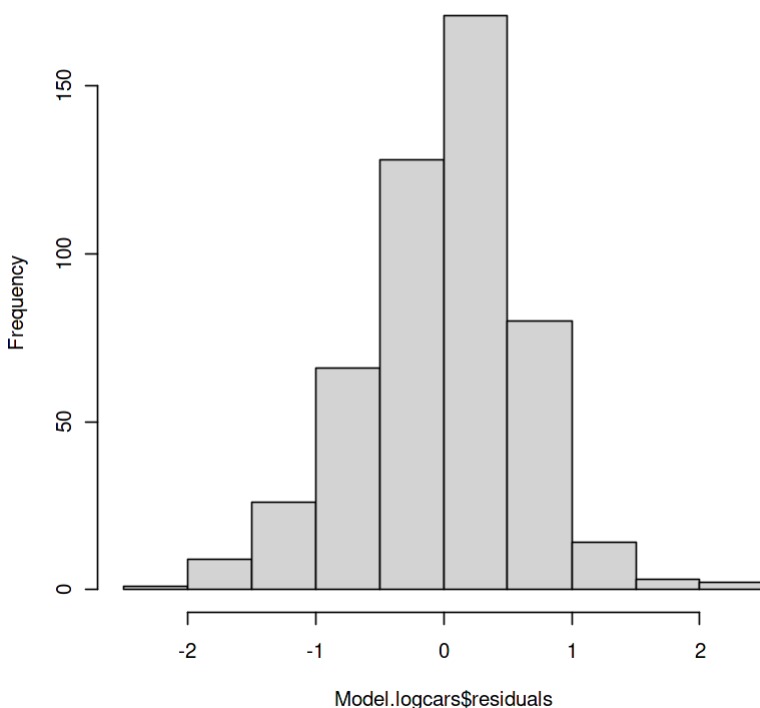


Fitted Values Of Model.temp
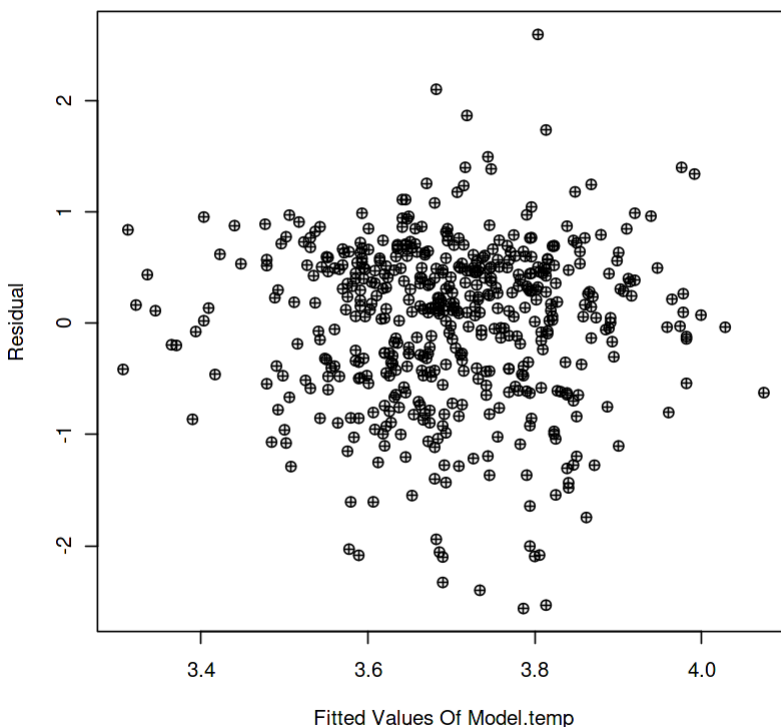
```
Call:
lm(formula = logNO2 ~ winds, data = z)

Residuals:
     Min       1Q   Median       3Q      Max
-2.35783 -0.43975  0.09252  0.49609  2.38463

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.12113    0.06299   65.43  < 2e-16 ***
winds       -0.13834    0.01780   -7.77 4.51e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7096 on 498 degrees of freedom
Multiple R-squared:  0.1081,    Adjusted R-squared:  0.1063
F-statistic: 60.38 on 1 and 498 DF,  p-value: 4.509e-14
```

**Histogram of Model.temp$residuals**

Model.temp$residuals

**Residuals For Model.winds**



Fitted Values Of Model.winds
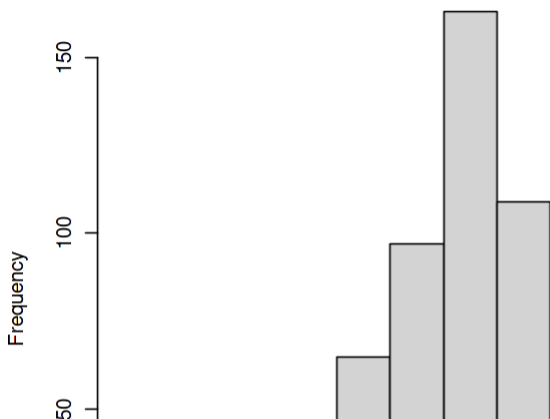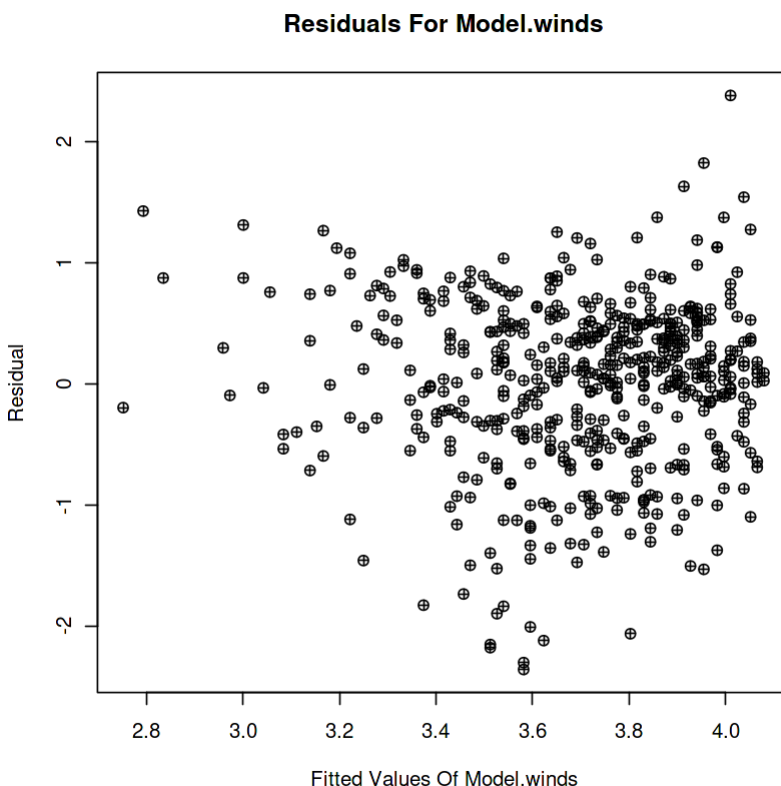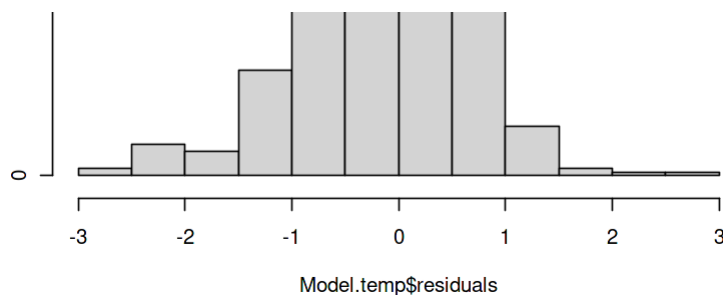
```
Call:
lm(formula = logNO2 ~ tempd, data = z)

Residuals:
    Min      1Q  Median      3Q     Max
-2.4931 -0.4407  0.1478  0.5050  2.5555

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.68003    0.03341 110.140  < 2e-16 ***
tempd        0.12273    0.03109   3.947 9.04e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7399 on 498 degrees of freedom
Multiple R-squared:  0.03034,   Adjusted R-squared:  0.02839
F-statistic: 15.58 on 1 and 498 DF,  p-value: 9.04e-05
```

**Histogram of Model.winds$residuals**
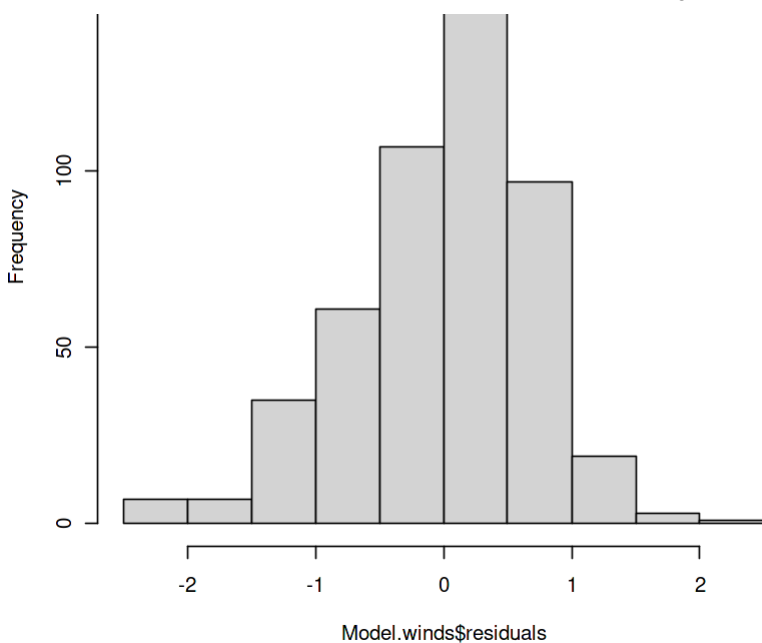
**Residuals For Model.tempd**
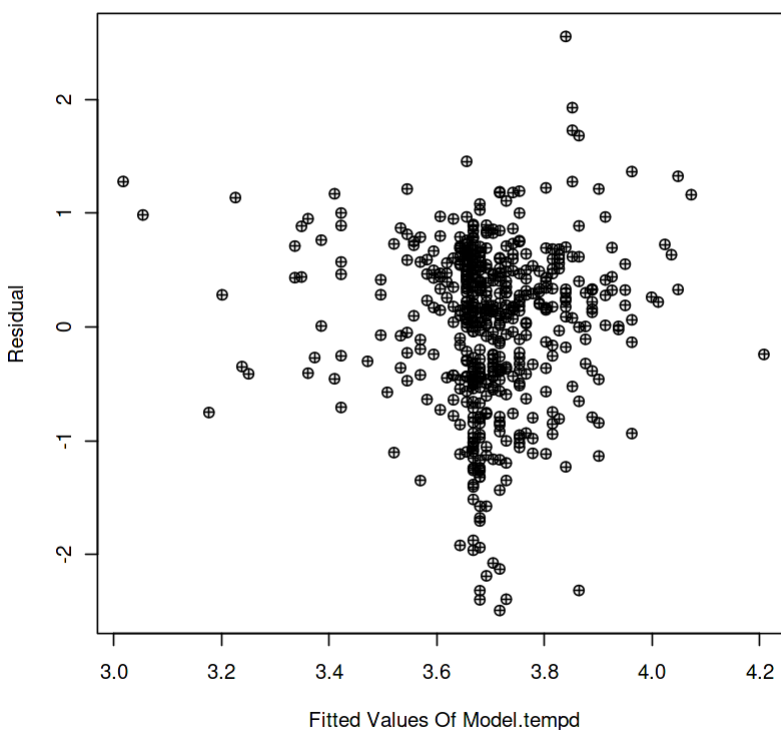


```
Call:
lm(formula = logNO2 ~ windd, data = z)

Residuals:
    Min      1Q  Median      3Q     Max
-2.3641 -0.4642  0.1442  0.5128  2.7519

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 3.5574971  0.0646677  55.012   <2e-16 ***
windd       0.0009826  0.0003863   2.544   0.0113 *
---
Signif. codes:   0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Signif. codes:   0        0.001      0.01     0.05   .  0.1       1
```

```
Residual standard error: 0.7465 on 498 degrees of freedom
Multiple R-squared:  0.01282,    Adjusted R-squared:   0.01084
F-statistic:  6.47 on 1 and 498 DF,  p-value: 0.01127
```

**Histogram of Model.tempd$residuals**
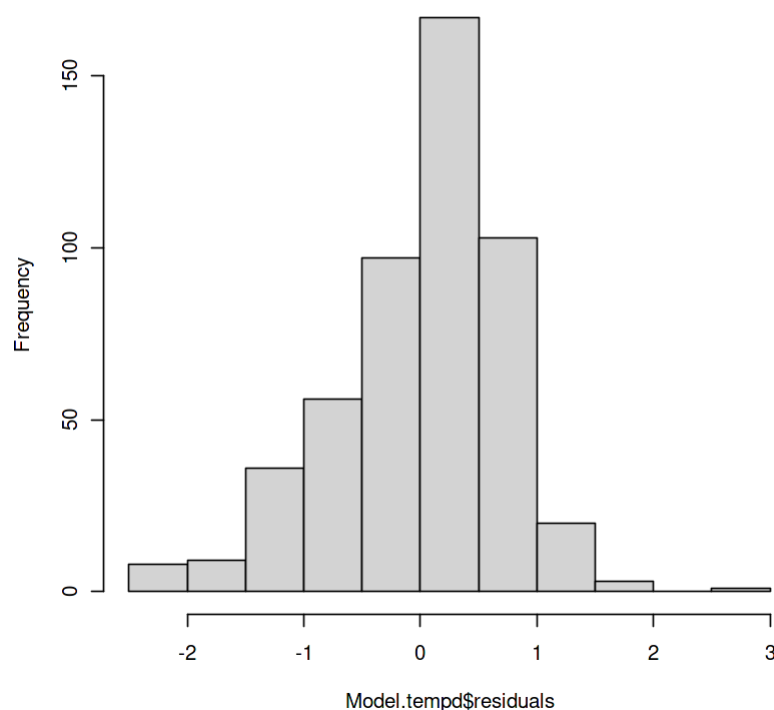


**Residuals For Model.windd**
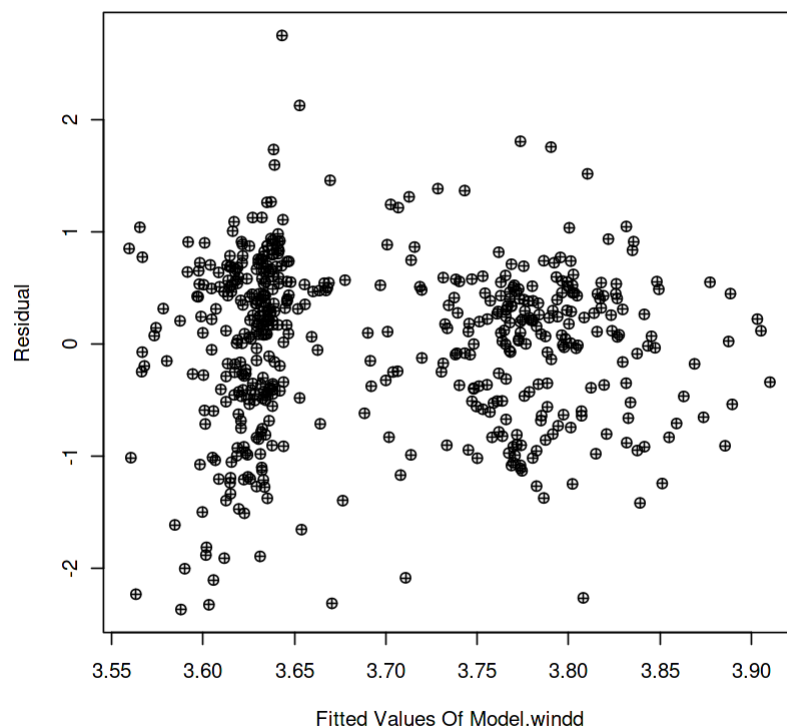


```
Call:
lm(formula = logNO2 ~ hour, data = z)
```

```
Residuals:
    Min      1Q  Median      3Q     Max
-2.2712 -0.4956  0.1000  0.5008  2.6528

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 3.362006   0.067687  49.670  < 2e-16 ***
hour        0.027165   0.004792   5.669 2.44e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7282 on 498 degrees of freedom
Multiple R-squared:  0.06062,   Adjusted R-squared:  0.05873
F-statistic: 32.13 on 1 and 498 DF,  p-value: 2.44e-08
```

**Histogram of Model.windd$residuals**



**Residuals For Model.hour**
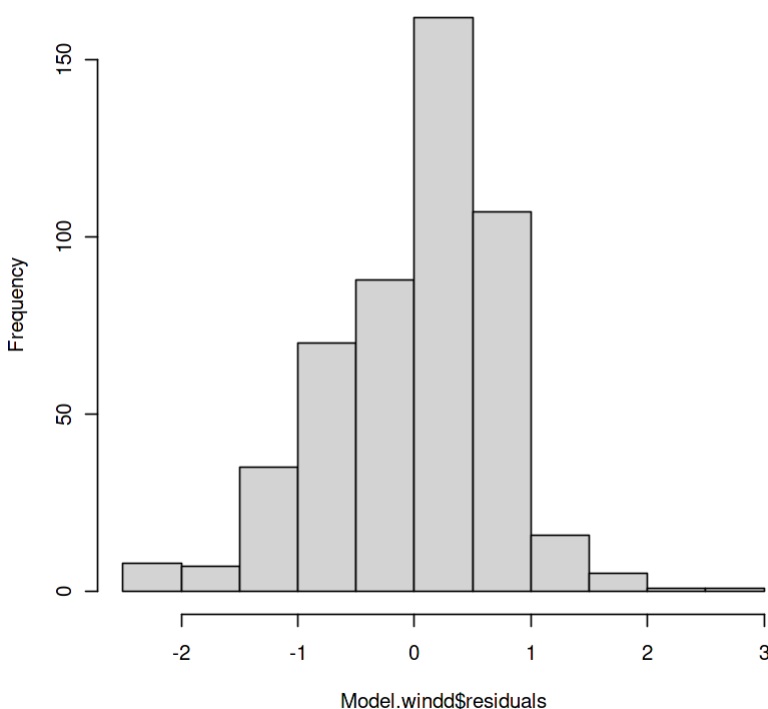
Fitted Values Of Model.hour
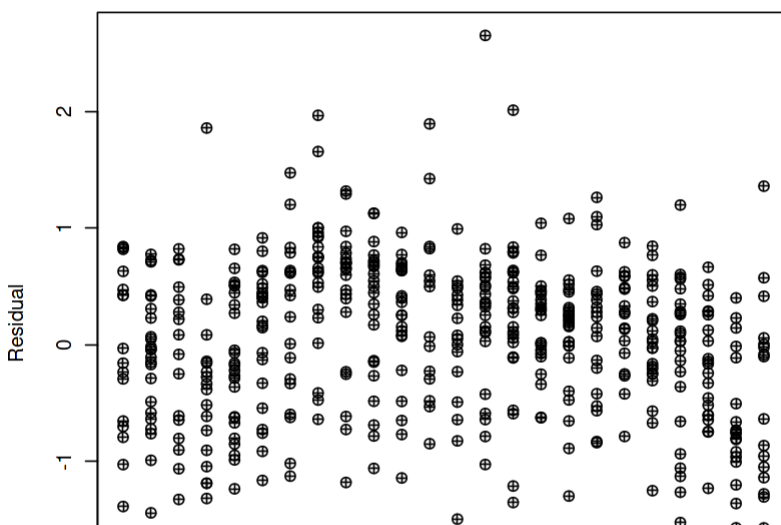
```
Call:
lm(formula = logNO2 ~ wind1, data = z)

Residuals:
    Min      1Q  Median      3Q     Max
-2.4121 -0.4581  0.1566  0.5092  2.7592

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.63592    0.04545   79.99   <2e-16 ***
wind1        0.13635    0.06716    2.03   0.0429 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7483 on 498 degrees of freedom
Multiple R-squared:  0.008207,   Adjusted R-squared:  0.006216
F-statistic: 4.121 on 1 and 498 DF,  p-value: 0.04288
```

### Histogram of Model.hour$residuals



Model.hour$residuals

### Residuals For Model.wind1
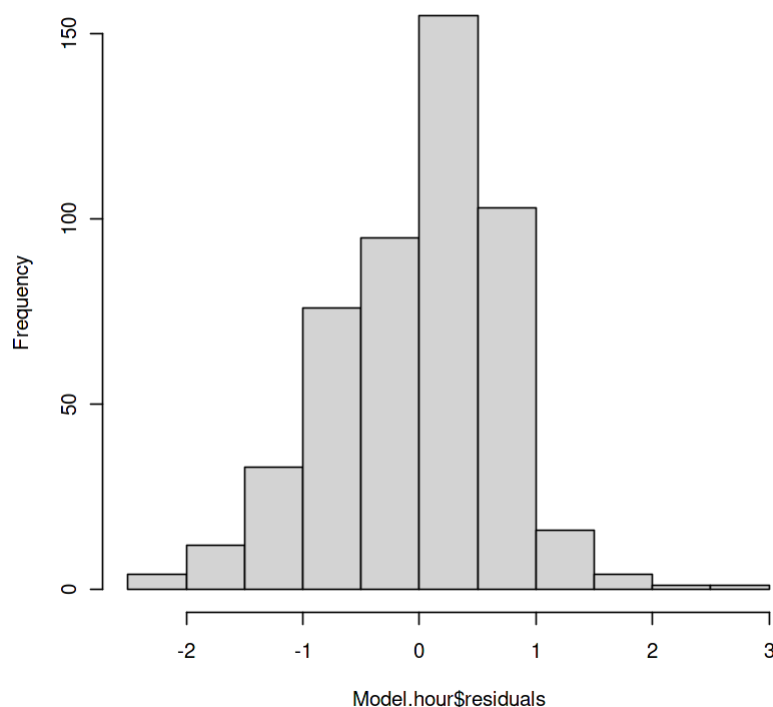
Fitted Values Of Model.wind1
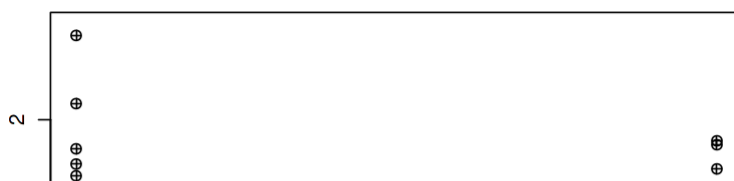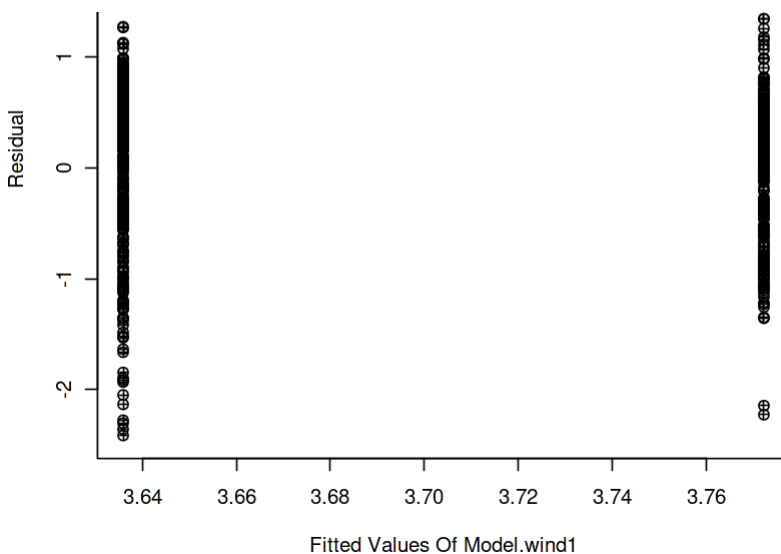
```
Call:
lm(formula = logNO2 ~ daytime, data = z)

Residuals:
    Min      1Q  Median      3Q     Max
-2.2554 -0.4276  0.1260  0.4577  2.4495

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.47917    0.04387  79.298  < 2e-16 ***
daytime      0.46639    0.06400   7.288 1.25e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7142 on 498 degrees of freedom
Multiple R-squared:  0.09637,   Adjusted R-squared:  0.09455
F-statistic: 53.11 on 1 and 498 DF,  p-value: 1.246e-12
```

**Histogram of Model.wind1$residuals**

Model.wind1$residuals

**Residuals For Model.daytime**
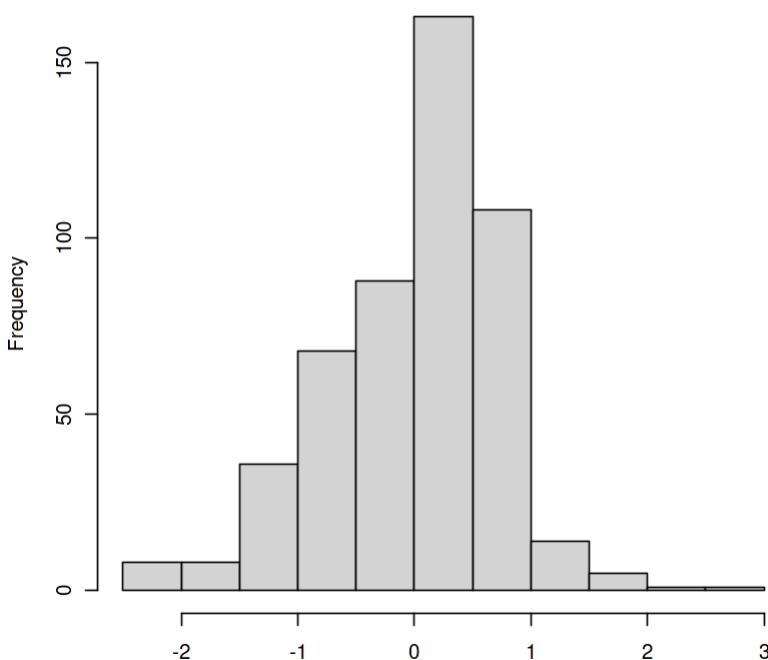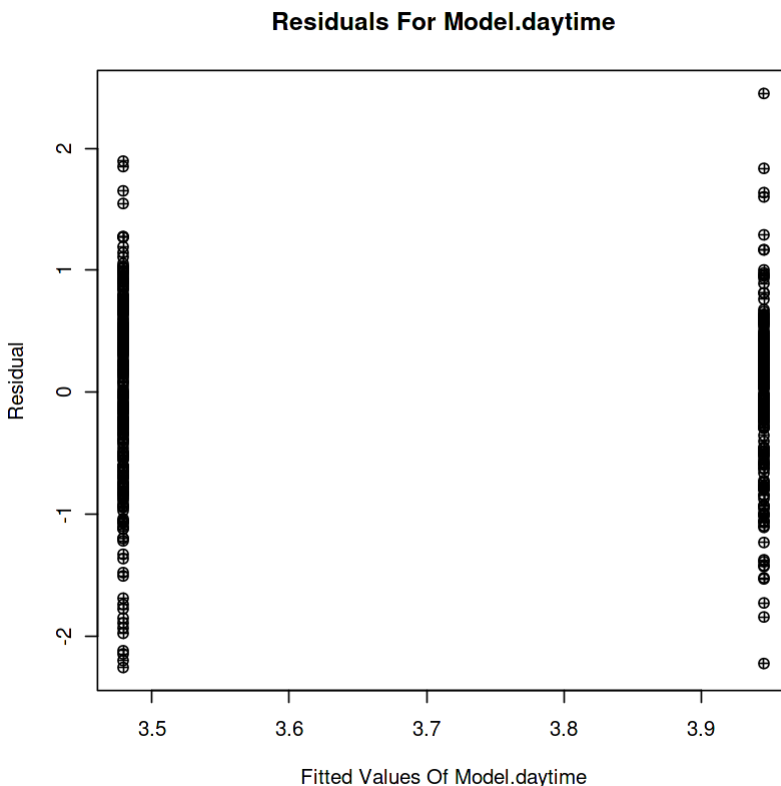


```
Call:
lm(formula = logNO2 ~ day7, data = z)

Residuals:
    Min      1Q  Median      3Q     Max
-2.4009 -0.4728  0.1293  0.5092  2.7212

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.62469    0.06096  59.463   <2e-16 ***
day7         0.02462    0.01701   1.447    0.148
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7498 on 498 degrees of freedom
Multiple R-squared:  0.004188,  Adjusted R-squared:  0.002189
F-statistic: 2.095 on 1 and 498 DF,  p-value: 0.1485
```

**Histogram of Model.daytime$residuals**

Model.daytime$residuals

**Residuals For Model.day7**
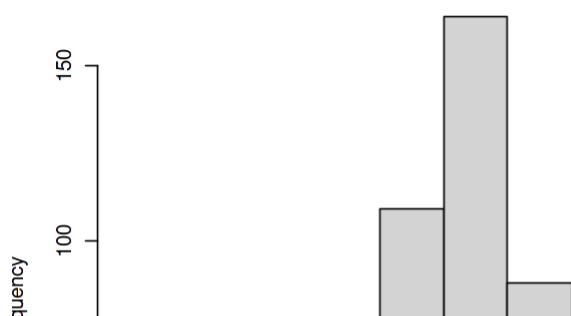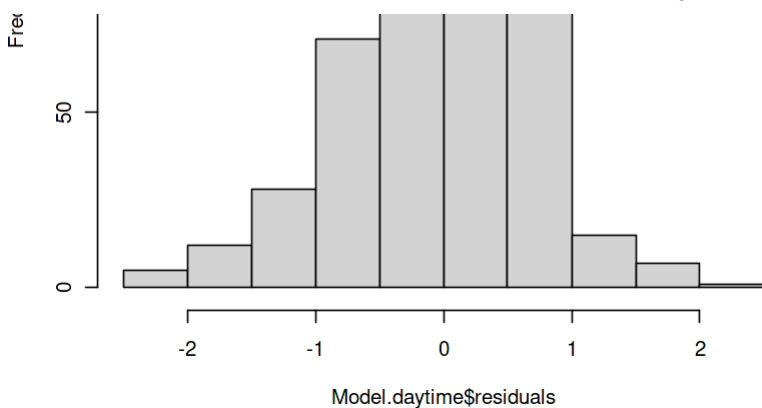


Fitted Values Of Model.day7
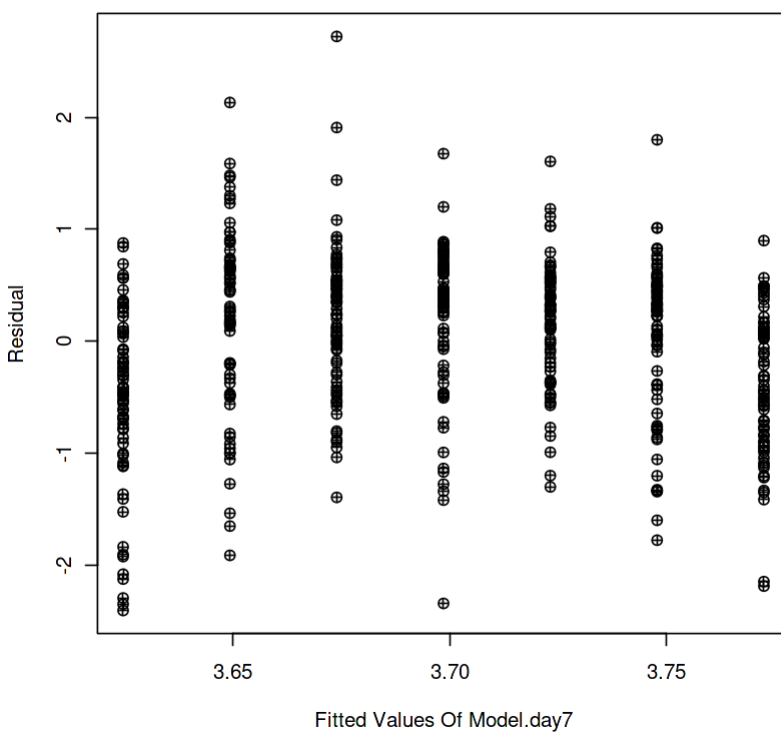
```
Call:
lm(formula = logNO2 ~ day365, data = z)

Residuals:
    Min      1Q  Median      3Q     Max
-2.5811 -0.4654  0.1440  0.5353  2.6606

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.8546855  0.0793241  48.594   <2e-16 ***
day365      -0.0012146  0.0005589  -2.173   0.0302 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7478 on 498 degrees of freedom
Multiple R-squared:  0.009394,  Adjusted R-squared:  0.007405
F-statistic: 4.723 on 1 and 498 DF,  p-value: 0.03024
```
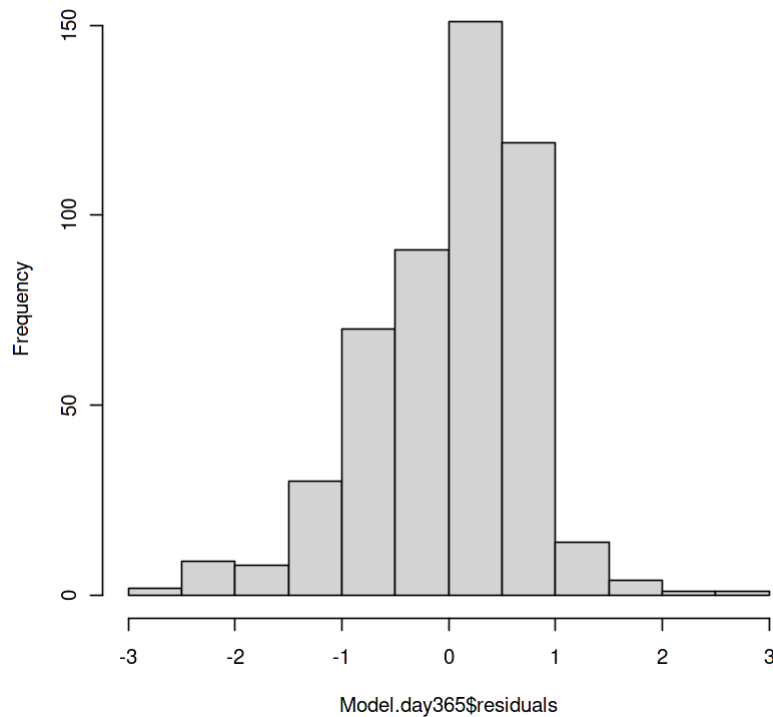
**Histogram of Model.day7$residuals**
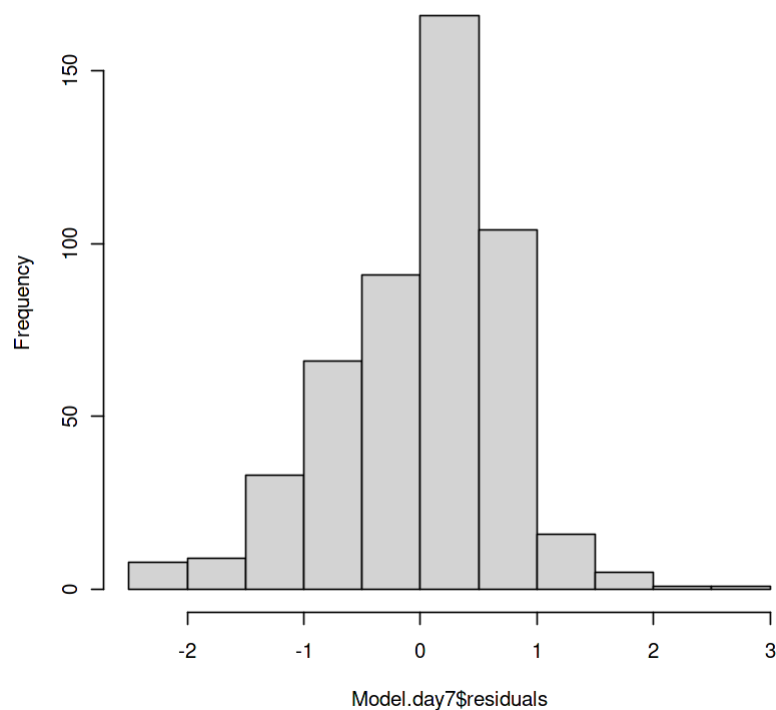
**Residuals For Model.day365**
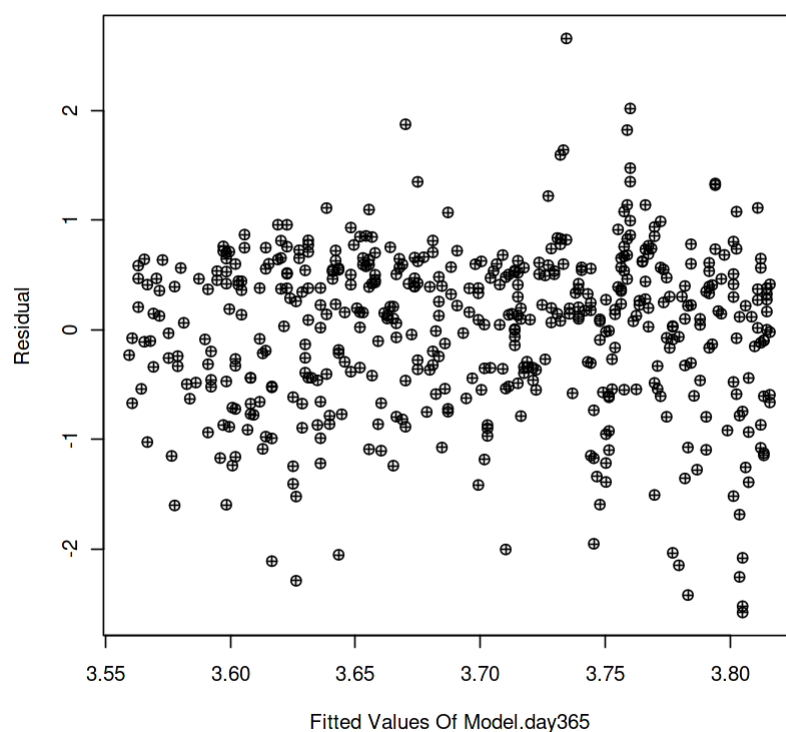


```
Call:
lm(formula = logNO2 ~ daysplit, data = z)

Residuals:
    Min      1Q  Median      3Q     Max
-2.5008 -0.4554  0.1401  0.4545  2.5333

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept)   3.86178     0.04222   91.459   < 2e-16 ***
daysplit     -0.39857     0.06594   -6.044 2.95e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7252 on 498 degrees of freedom
Multiple R-squared:  0.06834,   Adjusted R-squared:  0.06647
F-statistic: 36.53 on 1 and 498 DF,  p-value: 2.947e-09
```
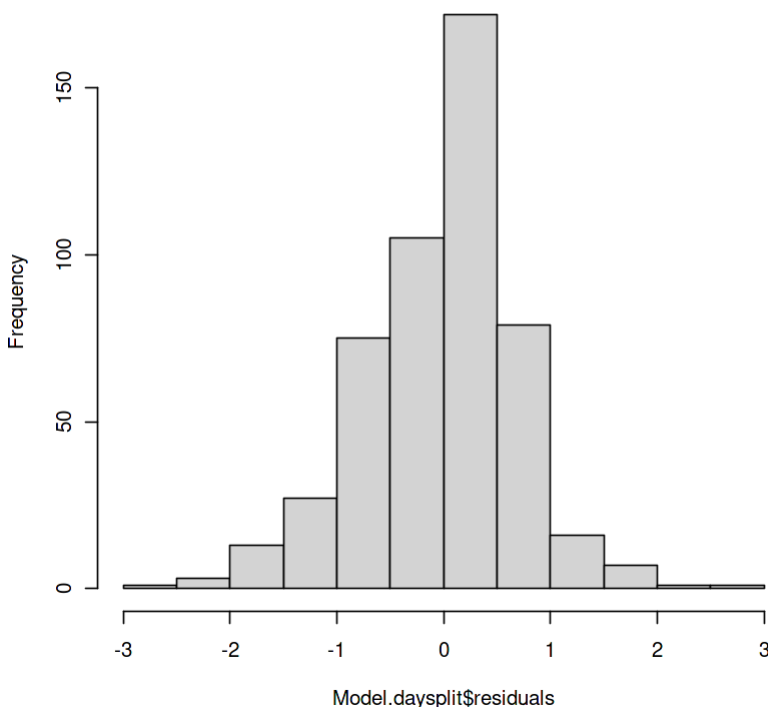
**Histogram of Model.day365$residuals**



**Residuals For Model.daysplit**
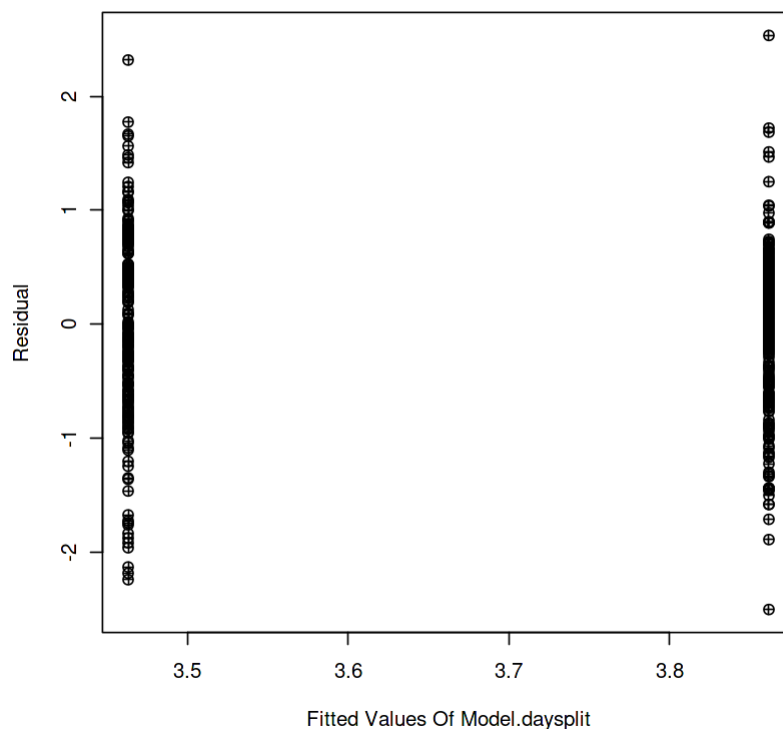
```
Call:
lm(formula = logNO2 ~ logcars + temp + winds + hour + daytime +
    daysplit, data = z)

Residuals:
     Min       1Q   Median       3Q      Max
-2.11147 -0.31127 -0.00571  0.36681  1.87029

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.299074   0.236347   5.496 6.23e-08 ***
logcars      0.457403   0.041533  11.013  < 2e-16 ***
temp        -0.026084   0.003779  -6.902 1.59e-11 ***
winds       -0.146978   0.013763 -10.679  < 2e-16 ***
hour        -0.014915   0.004884  -3.054  0.00238 **
daytime     -0.031791   0.074177  -0.429  0.66841
daysplit    -0.291323   0.048972  -5.949 5.13e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5327 on 493 degrees of freedom
Multiple R-squared:  0.5024,    Adjusted R-squared:  0.4963
F-statistic: 82.94 on 6 and 493 DF,  p-value: < 2.2e-16
```
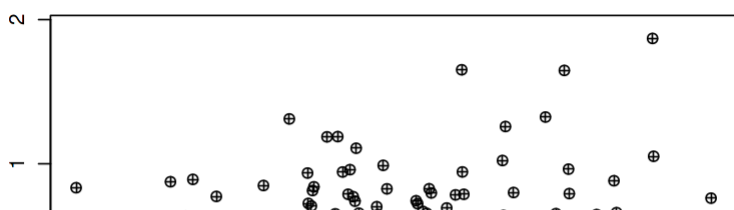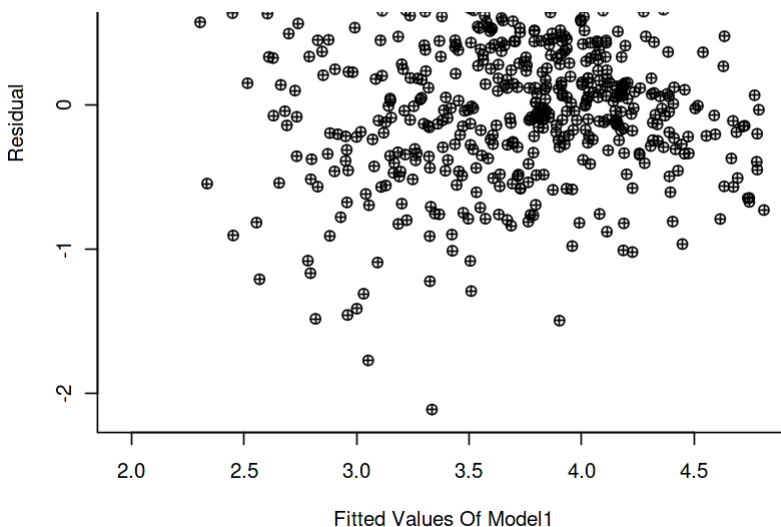
### Histogram of Model.daysplit$residuals



### Residuals For Model1

Fitted Values Of Model1

```
Call:
lm(formula = logNO2 ~ logcars + temp + winds + hour + daysplit,
    data = z)

Residuals:
     Min       1Q    Median       3Q       Max
-2.09434  -0.30500  -0.00121   0.36370   1.85733

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.368052   0.172938   7.911 1.69e-14 ***
logcars      0.444086   0.027539  16.126  < 2e-16 ***
temp        -0.026068   0.003776  -6.903 1.57e-11 ***
winds       -0.147835   0.013605 -10.866  < 2e-16 ***
hour        -0.013929   0.004305  -3.236  0.00129 **
daysplit    -0.292934   0.048788  -6.004 3.73e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5323 on 494 degrees of freedom
Multiple R-squared:  0.5022,    Adjusted R-squared:  0.4971
F-statistic: 99.66 on 5 and 494 DF,  p-value: < 2.2e-16
```
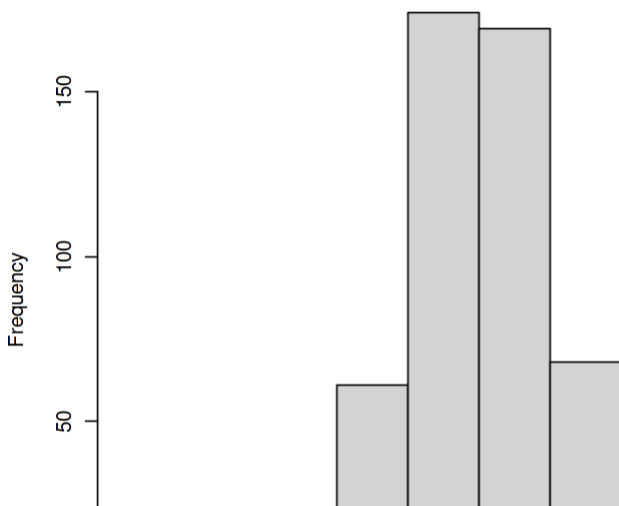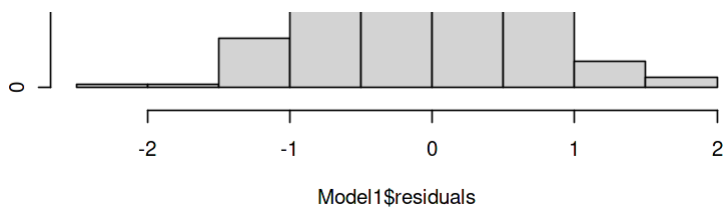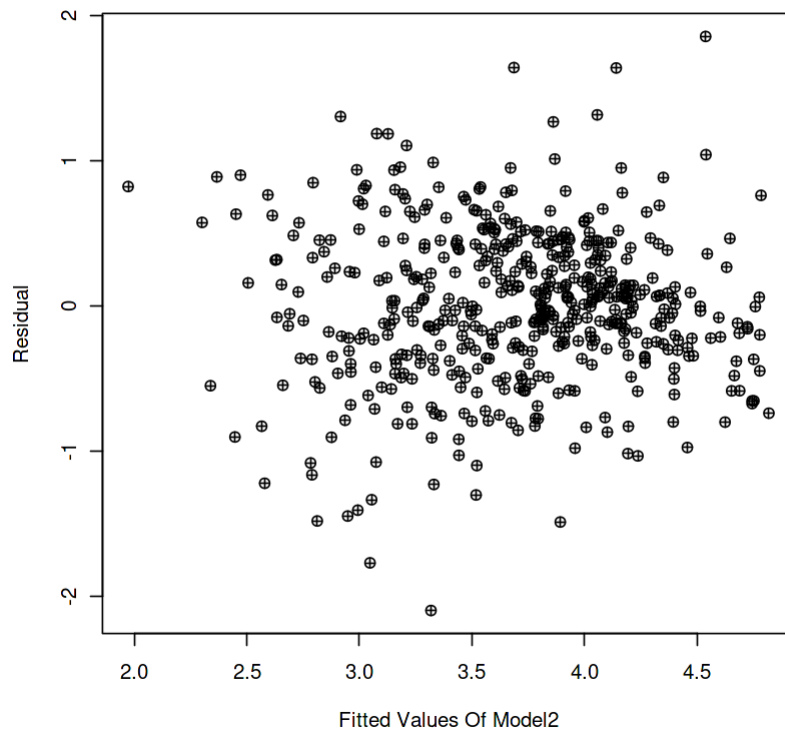
### Histogram of Model1$residuals

Model1$residuals

**Residuals For Model2**



Fitted Values Of Model2

**Histogram of Model2$residuals**



Model2$residuals